

**Exploring the role of Self/Other Perspective-Shifting in Theory of Mind  
with Behavioural and EEG Measures**

Elisabeth E.F. Bradford\*, Juan-Carlos Gomez, and Ines Jentzsch

University of St Andrews, UK

Word Count: 7,290 (main text)

Author Note

Correspondence concerning this article should be addressed to Elisabeth Bradford: University of Kent, School of Psychology, Keynes College, Canterbury, Kent, CT2 7NP, or Juan-Carlos Gomez, or Ines Jentzsch: School of Psychology & Neuroscience, University of St Andrews, St Andrews, Scotland, KY16 9JP.

Email: [E.E.Bradford@kent.ac.uk](mailto:E.E.Bradford@kent.ac.uk) (Bradford), [jg5@st-andrews.ac.uk](mailto:jg5@st-andrews.ac.uk) (Gomez), or [ij7@st-andrews.ac.uk](mailto:ij7@st-andrews.ac.uk) (Jentzsch).

\*Present address for Elisabeth Bradford: University of Kent, School of Psychology, Keynes College, Canterbury, CT2 7NP. Tel: 01227 824 240

Theory of Mind (ToM) refers to the ability to compute and attribute mental states to oneself and other people. This study sought to assess the extent of differentiation between 'Self' and 'Other' in ToM processes, and, of particular importance, the key role of perspective-shifting between 'Self' and 'Other'. Utilizing a newly established false-belief paradigm in a matched design, healthy adult participants completed the task whilst behavioural measures (response times, error rates) and electrophysiological (EEG) recordings were taken. Results revealed that self-oriented belief-attribution was faster and less error-prone than other-oriented belief-attribution, and demonstrated a key role of perspective-shifting. Perspective shifts from Self-to-Other resulted in longer response times and more errors than shifts from Other-to-Self. In contrast, no difference between self and other probes was found in no perspective-shift trials. Reflecting this, EEG recordings showed a significant interaction between Perspective-Shifting and Probe Type at an early onset across right parieto/occipito-lateral areas (250 ms post-stimulus onset), and across frontal-central areas from 500 ms post-stimulus onset, indicating the key role of these areas in ToM engagement. Results demonstrate that 'Self' and 'Other' can be distinguished at a behavioural level, and highlight the critical role of 'Perspective-Shifting' in ToM processes.

**Keywords:** theory of mind; social cognition; false-belief; perspective-shifting; belief-attribution

Everyday interaction and communication between individuals involves computation and attribution of various mental states, such as beliefs, desires, and intentions (Baron-Cohen et al., 2001). These abilities, known as ‘Theory of Mind’ (ToM; Premack & Woodruff, 1978), allow us to understand and negotiate everyday life in an appropriate manner. Importantly, ToM capacities apply not only to understanding other people’s mental states, but also one’s own mental states (Baron-Cohen et al., 2001). The importance of ToM abilities is illustrated in cases of Autistic Spectrum Conditions (ASC), where atypical or reduced ToM capacities are often cited as a defining feature (Baron-Cohen et al., 2001; Brent et al., 2004; although see also Gallese, Keysers, & Rizzolatti, 2004, and Gallese, 2006 for a counter-argument). In this paper, we examine, firstly, the extent to which ‘Self’ and ‘Other’ processes can be distinguished within the ToM mechanism and, secondly, the role of ‘Perspective-Shifting’ in ToM processes.

Many past studies examining ToM have focused on how individuals may compute and understand the mental states of other people (e.g. Gopnik & Astington, 1988; Wimmer & Perner, 1983). More recently, there has been additional focus on how individuals reflect on and understand their *own* mental states (e.g. Decety & Sommerville, 2003; Hartwright, Apperly & Hansen, 2012). A key question arising from this research is whether the same neural mechanisms are involved in both ‘Self’ and ‘Other’ mental state attribution, or whether these are two separate and distinct capacities, each relying on different cognitive components (Happé, 2003). Vogeley et al. (2001) comment that the extent to which ‘Self’ and ‘Other’ overlap in ToM is currently unclear, although they suggest that understanding of the ‘Self’ may be a necessary prerequisite to allow understanding of another individual’s mind; this suggestion is further supported by Epley, Keysar, Van Boven, & Gilovich (2004), who argue that in order to adopt another person’s perspective, individuals initially use the egocentric perspective as an ‘anchor’, adjusting their perspective-taking accordingly to allow consideration of the perspective of another person. Gallagher and Zahavi (2012) comment that it remains an ongoing debate as to whether understanding of the ‘self’ is a prerequisite for understanding of other people, suggesting distinct mechanisms, or whether ‘self’ and ‘other’

understanding is one-and-the-same, utilizing the same cognitive mechanisms. Further, it is as yet to be ascertained how it is that individuals switch between perspectives of 'Self' and 'Other'; is this an automatic, implicit process, or is it costly, requiring extra cognitive effort? (e.g. Saxe et al., 2006; McCleery et al., 2011).

One of the first tasks assessing differentiation between 'Self' and 'Other' in ToM was an unexpected contents task – the 'Smarties' task (Hogrefe, Wimmer & Perner, 1986). In the task, children aged 3-5 years are asked what they think is inside a sweets tube. On replying with 'sweeties'/'chocolate', children are shown that the tube actually contains a pencil. With the pencil then re-hidden, children are asked what they had thought to be in the tube, before seeing inside (self-oriented question), and additionally what a friend, who has not seen inside the tube, would think was inside (other-oriented). From 4-years-old, children reliably and consistently pass both the Self and Other oriented questions, correctly noting that they (and their friend) believed there to be sweets in the tube, even though they now know it contains a pencil. However, prior to 4-years-old, children tend to not only misrepresent the mental states of their friend (*'My friend would think there was a pencil inside'*) but also their own previous mental states (*'I thought there was a pencil'*). These results suggest that the ability to identify the mental states of both oneself and other people develop at the same time (i.e., our own minds are not understood first, leading to understanding of other people's minds, or vice versa). This could imply that mental state attribution recruits the same cognitive mechanisms, regardless of whether considering the 'Self' or 'Other'.

An issue with the Smarties task, however, was that the 'Self' question could potentially be passed without false-belief competence, with children simply recalling what they had previously verbally identified as the contents of the tube ('sweeties'), before they were shown the pencil (Williams & Happé, 2009). To address this, Williams and Happé (2009) conducted a revised unexpected contents task, the 'Plasters' task, in which explicit verbalization of the mental states was removed; children were asked to pass the experimenter a plaster, and in selecting the box they believed to contain the

plaster (from a choice of three boxes, only one of which was correct), a belief-state was established without requiring verbalisation. Whilst results were consistent for typically developing children, with equal performance on Self/Other-oriented trials, results for children with ASC showed that they performed significantly less well on *Self*-oriented questions than on Other-oriented test questions. These results indicate that individuals with ASC have greater difficulty representing their *own* beliefs than the beliefs of other people. Results such as these suggest that ‘Self’ versus ‘Other’ mental state attribution may rely upon two separate, albeit likely overlapping, neural mechanisms.

Further supporting this differentiation, recent studies on healthy adults’ ToM abilities have suggested that individuals adopt an egocentric view of the world, failing to consider the perspective of another individual unless explicitly motivated to do so (e.g., Cane, Ferguson, & Apperly, 2017; Epley et al., 2004; Ferguson et al., 2015; Keysar et al., 2000). In contrast, Samson et al. (2010) suggest that the mere presence of a secondary agent can influence adult’s performance on a given task; for instance, using a visual perspective taking paradigm, Samson and colleagues showed that participants were slower to report the number of dots visible in a room when a secondary agent (an avatar) could not see all the dots that the participant could see. Samson et al. suggest that these results indicate that participants are tracking the other agent’s perspective automatically, despite it not being task relevant, maintaining an awareness that others may have different knowledge states from the ‘self’ perspective. However, it is noted that there is much debate in literature regarding whether the visual perspective taking paradigm is measuring ‘social’ perspective-taking per se, or perhaps is driven by more domain-general attentional cues, such as directional cues of the avatar (e.g., Santiesteban et al., 2014; Catmur et al., 2016; Conway et al., 2017). Given these mixed findings, it currently remains unclear to what extent there is a differentiation between ‘Self’ and ‘Other’ in mentalising processes, and, crucially, how it is that an individual may alternate between each of these perspectives. The ability to smoothly switch between one’s own and other people’s perspectives – or to hold both egocentric and altercentric perspectives in mind concurrently (e.g., Samson et al., 2010) – plays a critical role in successful communication, allowing understanding of

another person's viewpoint, their current knowledge states, and how these may compare to one's own current state.

The present study attempts to explore how efficiently individuals are able to consider their own and other people's mental states using a computerised belief-attribution task. Participants are asked to answer simple dilemma questions from either their own or another person's perspective (e.g. '*You are/Joe is* looking for some sugar. Where would [you/Joe] look?'), before a true (sugar in the sugar jar) or false (marbles in the sugar jar) belief-state outcome is revealed. Participants are then asked to attribute a belief (probe stage), again from either their own or another person's perspective (e.g. '*What did [you/Joe] think was in the jar before seeing inside?*'). In some trials, the perspective participants are asked to respond from (Self/Other) changes across the dilemma and probe stages, thus creating a perspective-shifting scenario (Self-to-Other and Other-to-Self). These trials allow us to examine the extent to which being asked to respond from either the Self or Other perspective can influence outcomes, and also, of particular interest, the impact of a requirement to shift perspective from Self-to-Other or Other-to-Self.

A recent study (Bradford, Jentzsch, & Gomez, 2015), using this Self/Other Differentiation Task, revealed an interesting and surprising effect of perspective-shifting; results demonstrated that when shifting from Self-to-Other across a trial, performance was slower and less accurate than when shifting from Other-to-Self. In contrast, when no perspective-shift was required within a trial (Self-to-Self/Other-to-Other), there was no difference in performance (response time/error rates) between trial types. It is suggested that these results may reflect the role of the 'Self' in acting as the stem for understanding the 'Other'; the self-perspective is always processed, regardless of ultimate task demands, whereas the other-perspective is only processed when explicitly required. Thus, it would be more costly to switch from Self-to-Other, as the 'other' would not have been considered until required at the final stage of a trial, than it would be to switch from Other-to-Self, where the 'self' may have been processed as a means of understanding the 'other', thus simply requiring recall

of a previous belief-state (Bradford et al., 2015). This effect seems stable and is present across cultures, as recently shown in a cross-cultural study comparing Western and Chinese participants (Bradford et al., 2018). Furthering these findings, a recent study by Ferguson et al. (2017) used a level 1 visual perspective taking task (akin to Samson et al's (2010) paradigm) to assess the role of perspective shifting in this paradigm; results showed that, whilst a requirement to shift between the 'self' and 'other' perspective did not influence accuracy, participants were faster to respond in trials in which no perspective-shift was required compared to trials that did require a perspective-shift, suggesting that perspective-shifting itself is cognitively effortful.

The aim of the current paper was to further explore this perspective-shifting effect using the Self/Other Differentiation Task, further examining the differentiation between 'Self' and 'Other' within the ToM mechanisms, whilst seeking to investigate the timing and locations of the neurocognitive underpinnings of ToM processes using electrophysiological (EEG) measures. By using EEG measures, we sought to identify some of the neural correlates that may be supporting social cognition abilities and to identify temporal aspects of ToM processing in the brain. Amodio and Frith (2006) comment that there is currently only limited research that has studied the neural substrates underlying social cognition, and in focusing on methodologies assessing this, we may be better able to identify the unique, shared and coordinated roles of various neural structures in successful ToM engagement, particularly in healthily functioning adults.

There are currently relatively few studies that have used Event Related Potentials (ERPs) to study ToM in the brain, with some key exceptions; for instance, Meinhardt et al. (2011) conducted a study exploring true- and false-belief reasoning using ERPs. In their study, children and adult participants were presented with cartoon stories in which a true and false-belief scenario was created. Results revealed both a late positive complex and late anterior slow wave, which Meinhardt et al. suggested were associated with the neural responses of the 'decoupling mechanism' required for false-belief understanding. These results support the findings of Liu, Sabbagh, Gehring, & Wellman's (2009)

study in which adults and children completed a task in which they were required to reason about a character's beliefs in a series of animated vignettes; results demonstrated a frontal late slow wave that was only seen in participants who were able to correctly engage in false-belief reasoning, compared to participants who failed the false-belief reasoning task. Finally, Sabbagh et al.(2009) conducted a study in which pre-school children completed a battery of representational ToM tasks whilst EEG data was recorded; results showed that individual differences in the dorsal medial prefrontal cortex (dMPFC) and the right temporal-parietal junction (rTPJ) were positively associated with ToM task performance, suggesting a critical role in ToM development. However, whilst these prior studies have revealed differences in how true- and false-belief state scenarios may be processed, in the current study, we sought to explore potential differentiations between how 'self-oriented' or 'other-oriented' belief-states are processed.

Beyond studies using EEG/ERP measures, there have been some previous attempts to locate ToM in the brain using functional brain-imaging measures, and results have revealed various suggestions for where ToM processing may occur, including the medial prefrontal cortex (mPFC), rostral anterior cingulate cortex, medial posterior parietal cortex (mPPC), and the bilateral temporal parietal junction (TPJ) (Amodio & Frith, 2006; Happé, 2003; Mahy, Moses & Pfeifer, 2014; Mitchell, Macrae, & Banaji, 2006; Saxe & Wexler, 2005; Gallagher et al., 2000; Schurz et al., 2014; Saxe & Kanwisher, 2003). A meta-analysis of functional brain-imaging studies by Schurz et al. (2014) suggested a core ToM network that involves both the mPFC and the TPJ. Amodio and Frith (2006) suggested a unique role of the medial frontal cortex in relation to social cognition, particularly in terms of differentiation between 'self' and 'other' and, even more specifically, differentiating between similar and dissimilar 'others', suggesting separate processing of these capacities as well. Some research has further attempted to identify the specific roles different brain regions may play in ToM processing; for instance, prior studies have suggested that parietal areas, including the temporal-parietal cortex (Happé, 2003; Schurz et al., 2014) or posterior cingulate cortex (Gobbini et al., 2007; McCleery et al., 2011), may be involved in the disengagement of the 'self' perspective and adoption of the 'other'



perspective. McCleery et al. (2011), using a visual-perspective taking task, suggested that the posterior cortex (specifically, the temporo-parietal cortex) is responsible for the calculation and representation of 'Self' vs. 'Other' perspectives, with the right frontal cortex later recruited to aid resolution of conflicting perspectives. It has also been suggested that frontal areas such as the anterior cingulate cortex (Gobbini et al., 2007) and the pre-frontal cortex (Saxe & Kanwisher, 2003; Gallagher et al., 2000) may play a role in managing, or reconciling, contrasting perspectives (Saxe et al., 2006; Samson et al., 2005). However, previous studies have often relied on either comparing neural activity across ToM and non-ToM paradigms (e.g. false-beliefs versus false-photographs tasks; Sabbagh & Taylor, 2000) or have focused on attribution of beliefs to other people rather than the self (e.g., Gobbini et al., 2007; Gallagher et al., 2000), reducing the opportunity to directly compare abilities across these conditions, and to explore the process of shifting between perspectives.

Given these prior findings, our ERP analysis will focus on selected key areas of interest – central parietal areas and frontal-central areas – as well as exploring the presence of potential hemispheric lateralized effects, across frontal-lateral and parieto/occipito-lateral areas (McCleery et al., 2011; Saxe & Kanwisher, 2003; Saxe & Powell, 2006; Saxe et al., 2009). Given the exploratory nature of our ERP analysis, these focus areas were chosen based upon suggestions from prior literature about the key roles that these broad areas play in ToM processes, in particular the role of representation and reconciliation of 'Self' and 'Other' perspectives (e.g. Amodio & Frith, 2006; Happé, 2003; Saxe et al., 2006; McCleery et al., 2011; Schurz et al., 2014; Gobbini et al., 2007). A key strength of the Self/Other Differentiation Task is the use of a matched-design, in which 'self' and 'other' oriented conditions, including perspective-shift trials, are matched to allow direct comparison of performance in each of these trial types. In utilizing the Self/Other Differentiation task, the current study sought to establish whether 'Self' and 'Other' belief-attribution processes could be distinguished in healthy adults, and the effect of shifting perspectives between 'Self' and 'Other', using behavioural and EEG measures that allow direct comparison of the neural and cognitive requirements, and time-course, of these different capacities within the ToM mechanism.

It is predicted that trials which require participants to adopt the 'Other' perspective will be significantly slower and less accurate than trials in which the 'Self' perspective is required, supporting the idea that the 'self' is processed first, or more automatically, before the 'other' perspective is considered – or even that these two capacities are entirely dissociable, independently relying on different brain regions, with 'other' processing more demanding than 'self' processing, requiring extra cognitive resources that may lead to longer and more error prone processing. It is hypothesised that trials in which Perspective-Shifting is required (Other-to-Self/Self-to-Other) will be more difficult than those in which no perspective-shift is required (Self-to-Self/Other-to-Other), requiring more cognitive effort to switch to a different perspective than to maintain a perspective in no-shift trials. Importantly, it would be expected that trials in which a participant is required to switch from Self-to-Other would be significantly slower and more error prone than trials in which they are required to switch from Other-to-Self, as, in the latter, the 'Self' will have been processed egocentrically, to act as a 'stem' for understanding the 'Other' perspective, whereas in the former, there will have been no need to engage in any extra cognitive processing until explicitly required to do so at the probe stage.

Our ERP analysis, based on prior research suggestions, will focus on central-parietal areas and frontal-central areas, as well as frontal-lateral and parieto/occipito-lateral areas to explore the presence of potential lateralized effects (e.g. Happé, 2003; McCleery et al., 2011; Saxe & Kanwisher, 2003; Gobbini et al., 2007; Geng & Vossel, 2013). It is expected that these key areas will demonstrate significant differences in amplitudes mirroring behavioural results; that is, a differentiation between 'Self' and 'Other' oriented trials, as well as the role of perspective-shifting in ToM abilities, reflecting differences in the processing of these different trial types. Due to the exploratory nature of this analysis, and limited prior research directly comparing 'Self' and 'Other' oriented processing or the role of perspective-shifting, whilst an explicit differentiation between these components was predicted in our ERP analysis, the direction of these effects was not hypothesized.

## METHOD

### *Participants*

Thirty-six adult participants (26 females, 10 males; mean age 22.1 years, range 17-51 years), who all either spoke English as their first language, or were fluent in English, took part in this study. Participants received a small honorarium of £5 per hour as compensation for time spent participating. One participant (male) was excluded due to a computer fault (failure to record data), thus thirty-five participants completed the study in total. To ensure homogeneity of the sample group for analysis, one further participant was excluded from analysis (female, 51 years). This meant that the total sample size for analysis was thirty-four adult participants (25 females, 9 males; mean age 21.03 years, range 17-30 years). All participants gave informed consent and this study was approved for use in human subjects in accordance with the University of St Andrews Research Ethics committee.

### *Apparatus*

Stimuli were presented on a 17-inch cathode ray tube (CRT) monitor. EEG data was recorded in a Faraday cage; the computer monitor was outside of the cage, visible through a window in the shielded booth, with a viewing distance of 80 cm. A keyboard was used to record responses, with the three arrow keys ( $\leftarrow \downarrow \rightarrow$ ) assigned as answer keys (the position of the key corresponded to the spatial location of answer stimuli – see ‘procedure’). Coloured images were presented on a white background, in the centre of the screen. Questions were presented in black text.

A BIOSEMI Active-Two amplifier system with 70 Ag/AgCl electrodes was used to record electroencephalographic (EEG) activity. Two additional electrodes, the Common Mode Sense (CMS) active electrode and Driven Right Leg (DRL) passive electrode, were used as reference and ground electrodes, respectively. Electrodes were mounted on an elastic cap. The two-electrode CMS/DRL feedback loop calculates and equates the average potential of each participant (the Common Mode

voltage) to the reference voltage of the Analog-to-Digital Conversion apparatus; EEG was acquired at the scalp electrode in 'raw' mode, relative to this reference voltage drawn from the CMS electrode, bringing the average potential of each participant as close as possible to the amplifier ground and reducing noise levels (Sysoeva, Lange, Sorokin & Campbell, 2015; Grimm, Roeber, Trujillo-Burreto, & Schröger, 2006; Campbell, Kerlin, Bishop, & Miller, 2012). The use of active electrodes removed the need for skin preparation (e.g. abrasion). Horizontal eye-movements were monitored by electrodes placed on the outer canthi of each eye, whilst vertical eye movements were monitored by electrodes placed beneath each eye. EEG recordings were sampled at 256 Hz. Off-line all EEG channels were recalculated to average reference. Trials containing blinks and horizontal eye movements were corrected using the adaptive artefact correction method of Brain Electromagnetic Source Analysis (BESA) software (Ille, Berg & Scherg, 2002).

### ***Procedure***

Participants were first prepared for EEG recording outside of the testing booth (electrode application). They were then taken into a testing booth, and seated in front of a computer screen in a dark room.

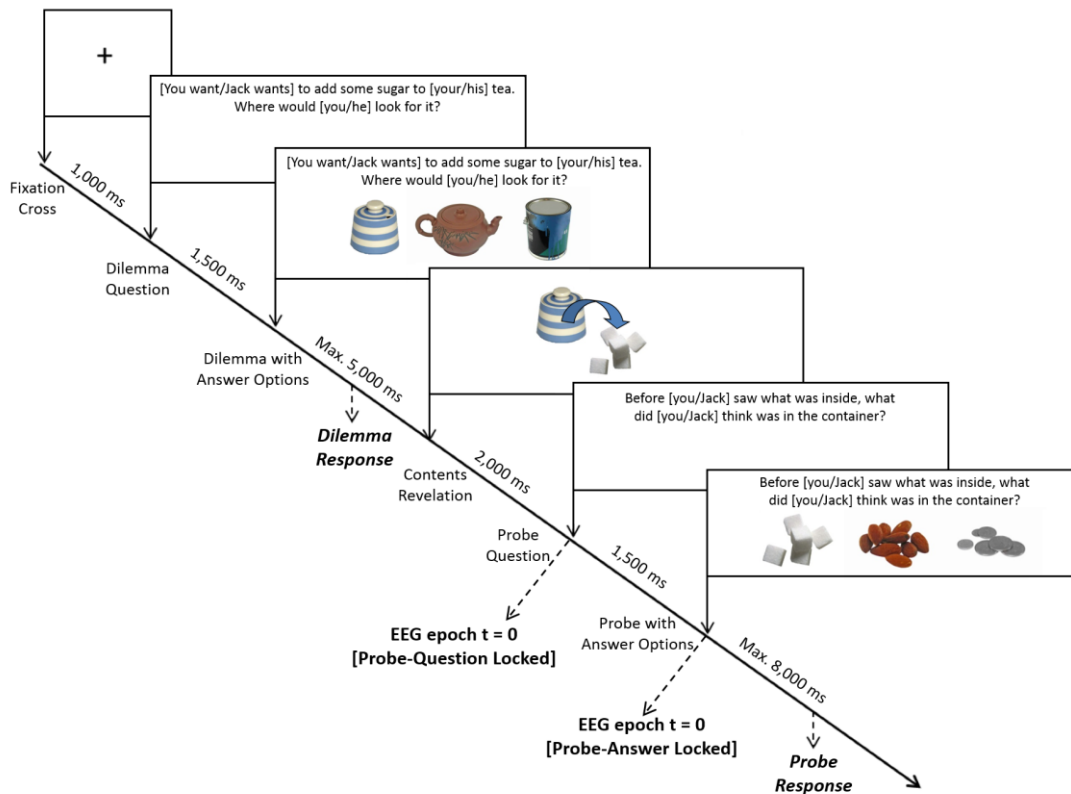
In contrast to our previous study (Bradford et al., 2015), here the Self/Other Differentiation Task was programmed in ERTS (Experimental Run Time System) rather than in E-Prime, and the number of experimental trials included was increased. The task consisted of 8 practice trials and 216 test trials, with each trial consisting of three stages: dilemma stage, contents revelation stage, and probe stage. Both the practice and test trials followed the same format (dilemma→contents→probe), but only test trials required belief-attribution. Table 1 provides details of trial types and the number of trials in each condition; Figure 1 provides an example of the time course of a single trial.

**Table 1:** Number of trials in each condition combination of the Self/Other Differentiation Task. Distracter trials refer to probe stage questions that do not require belief-attribution (e.g., ‘*What colour was the sugar bowl?*’) included to reduce participants anticipation of the correct answer in critical probe questions.

	Self Dilemma			Other Dilemma			Total
	Self Probe	Other Probe	Distracter	Self Probe	Other Probe	Distracter	
Expected Contents	17	17	20	17	17	20	108
Unexpected Contents	17	17	20	17	17	20	108
Total	34	34	40	34	34	40	216

*Dilemma Stage* – to create a belief state, participants were asked where they would look for a specific object, indicating their answer through selection of an appropriate container from three images presented in a horizontal line. Dilemmas could be either self-oriented (‘*You are looking for a biscuit to eat. Where would you look?*’) or other-oriented (‘*Jane is looking for a biscuit to eat. Where would she look?*’). In practice trials, participants were asked to select a specific object (e.g. ‘*Select the backpack*’). Dilemma questions were shown alone for 1500 ms, before image answer options were also displayed, for a maximum of 5000 ms. If participants selected an incorrect answer/failed to indicate an answer, an ‘X’ was displayed for 1500 ms before the correct answer was revealed at the contents revelation stage.

*Contents Revelation Stage* – contents of a container could be either expected (true belief) or unexpected (false belief). Contents were shown for 2000 ms and did not require a response. Following self/other oriented dilemmas, half of each were followed by expected contents, and half by unexpected contents. In practice trials, contents was always expected (e.g. folders in the backpack).



**Figure 1:** Schematic example of the presentation sequence across a single trial in the Self/Other Differentiation task, starting with the Dilemma Question, followed by Contents Revelation, and finally the Probe Question. For this example, the comparative Self/Other conditions are presented together (You want/Jack wants). Trials could either involve no perspective shift (Self-Self, Other-Other), or a perspective shift between the dilemma and probe stage (Self-Other, Other-Self).

*Probe Stage* – used to assess belief-attribution abilities, the probe could be either self-oriented or other-oriented. Participants were asked to identify what either they themselves or another person believed to be in the container, before they saw inside. Answers were indicated by selection of one image from three presented in a horizontal line. Distracter questions were also included at the probe stage (e.g. ‘What colour was the wallet?’) to reduce participant’s ability to anticipate the correct answer until after seeing the probe question. The probe question was displayed for 1500 ms before the three answer options were revealed. Image answer options were displayed until response, or for a maximum of 8000 ms, if no response was recorded. In practice trials, participants were asked to select the picture that showed what had been in the container (i.e. reality/true belief).

Response times were recorded at both dilemma and probe stages. ERP analysis focused on neural responses to the probe question. To create the 'Self' and 'Other' comparable conditions, sentences at both the dilemma and probe stages were matched in terms of syntax and syllables, so that each sentence had at least one paired match in the opposite condition (Self vs. Other). In the 'Other' conditions, all names used were one syllable, ensuring the length of sentences was not altered. A fixation cross (+) was displayed for 1000ms between trials.

### ***Analysis of Behavioural Responses***

Response times and error rates were assessed for responses to both the Dilemma Stage and Probe Stage of the Self/Other Differentiation Task, to assess differentiation in performance (i.e., speed and accuracy) between 'Self' and 'Other' oriented trials. Prior to analysis, data was screened for outliers in overall reaction times by removing response times that were outside of three standard deviations from the overall reaction time mean. At both the Dilemma and Probe Stage, response times were time-locked to the onset of the three answer image options (i.e., once a participant was able to provide a response). Responses to the dilemma question were analysed using a paired samples t-test, comparing responses to 'self' and 'other' oriented dilemmas. Responses to the probe question were analysed by taking into account whether there was a shift of perspective between Dilemma Type and Probe Type. For trials in which there was no perspective shift, both the Dilemma and Probe addressed either the self (Self-Self) or the other (Other-Other). For trials in which there was a perspective shift between Dilemma and Probe stages, the shift could be from either 'self' at the dilemma stage to 'other' at probe stage (Self-Other), or from 'other' to 'self' (Other-Self). To analyse the probe question responses, a 2 (Perspective Shift: No Shift vs. Shift) x 2 (Contents: Expected vs. Unexpected) x 2 (Probe: Self vs. Other) Repeated-Measures ANOVA was conducted.

### ***EEG Recording and Analysis***

Trials with incorrect, missing, too slow (over three standard deviations from overall reaction time mean), or too fast (below 100 ms) responses, or with EEG artefacts (e.g., amplifier blocking, scalp muscular activity, and slow linear drift) were excluded from the ERP data analysis. ERP data was analysed in epochs locked to the onset of the probe question (Question-locked ERPs) and the probe answer options (Answer-Option-locked ERPs) (see also Figure 1). Analysis epochs lasted for 1200ms, starting 200ms before the onset of the probe question (for Question-locked ERPs) or probe answer options (for Answer-locked ERPs).

### *Regions of Interest*

Four exploratory regions of interest analysis were defined, with focus electrode groups selected for analysis based on suggestions from prior research (e.g., Hartwright et al., 2012; McCleery et al., 2011; Schurz et al., 2014; Saxe & Kanwisher, 2003) and visual inspection of the topographic maps, which indicated the locus of the sites of maximal effects (see Figures 3 and 4). The two broad regions of interest selected were: fronto-central areas (electrodes F1, FZ, F2, FC1, FCz, and FC2) and parieto-central areas (electrodes CP1, P1, CPz, PZ, P2, and CP2). We were also interested in the lateralization of any potential effects, exploring potential hemisphere differences throughout effects; thus, analysis also looked at fronto-lateral areas (electrodes F5, FC5, F7, FT7, F6, FC6, F8, and FT8), and parieto/occipito-lateral areas (electrodes O1, PO3, PO7, P7, P5, P9, O2, PO4, PO8, P6, P8, and P10).

Values were measured as mean amplitudes in subsequent 50ms intervals, from 150ms to 750ms post-stimulus presentation, using a 200ms pre-stimulus baseline. Similar to the analysis of the behavioural data at probe stage, frontal-central and parietal-central ERP amplitudes were subjected to 2 (Perspective Shift: No Shift vs. Shift) x 2 (Contents: Expected vs. Unexpected) x 2 (Probe: Self vs. Other) Repeated-Measures ANOVAs. For analysis of frontal-lateral and parieto/occipito-lateral ERP amplitudes, an additional factor of hemisphere (left vs. right) was included.



To reduce the possibility of Type 1 errors in the comparisons in the analysis, we have constrained our analysis in two ways. First, given that our main research interest was to explore the neural correlates of the probe x perspective-shift effect, we will only report significant results that include an interaction between shift and probe. Second, we will only consider results as significant if for a given effect  $p$ -values smaller than 0.05 are found in at least three subsequent time windows. Table 2 presents  $F$  values for ERPs time-locked to the Probe Question onset, and Table 3 presents  $F$  values for ERPs time-locked to the Probe Answer-Options onset (see Figure 1).

## RESULTS

### ***Behavioural Data: Response Times (RT) and Error Rates***

For response times, only responses to correct trials were analysed. Accuracy was overall very high; for the dilemma stage, an average of 2.2% trials were removed per participant, whilst for the probe stage, an average of 3.9% trials were removed per participant. Response times were time-locked to the onset of the three answer option images.

#### *Dilemma Stage*

A paired-samples  $t$ -test revealed no significant difference in RT,  $t(33) = -1.72$ ,  $p = .10$ ,  $r = .24$ , or error rates,  $t(33) = .40$ ,  $p = .69$ ,  $r = .09$  between Self-Oriented and Other-Oriented dilemma questions,  $M(\text{RT}) = 1279$  vs. 1311 ms;  $M(\text{Error Rate}) = 2.29$  vs. 2.12 %.

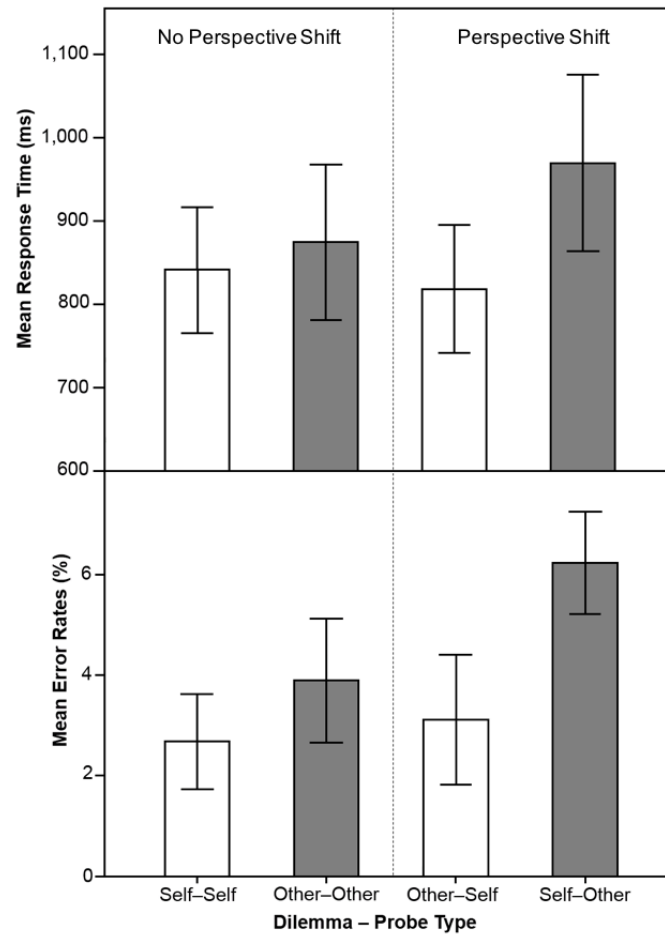
#### *Probe Stage*

A 2 (Perspective Shift: No Shift vs. Shift) x 2 (Contents: Expected vs. Unexpected) x 2 (Probe Type: Self vs. Other) Repeated-Measures ANOVA was conducted on RT data and error rates. Figure 2 depicts mean RTs and error rates in response to the probe question in the Self/Other Differentiation task.

Results revealed a significant main effect of Probe Type, with faster,  $F(1, 33) = 26.38, p < .001, \eta_p^2 = .44$ , and more accurate,  $F(1, 33) = 22.81, p < .001, \eta_p^2 = .41$ , responses to self-oriented probes (830 ms; 2.90% errors) than other-oriented probes (922 ms; 4.93% errors). A significant main effect of Perspective-Shifting was present, with faster,  $F(1, 33) = 5.07, p = .03, \eta_p^2 = .13$ , and more accurate,  $F(1, 33) = 15.06, p < .001, \eta_p^2 = .31$ , responses in no perspective-shift trials (859 ms; 3.16% errors) than in perspective-shift trials (892 ms; 4.67% errors). A main effect of Contents Type showed that responses were faster,  $F(1, 33) = 65.18, p < .001, \eta_p^2 = .66$ , and more accurate,  $F(1, 33) = 124.30, p < .001, \eta_p^2 = .79$ , following expected contents trials (824 ms; 1.34% errors) than unexpected contents trials (928 ms; 6.49% errors).

Importantly, and replicating previous behavioural results (Bradford et al., 2015), there was a significant interaction between Perspective Shift and Probe Type, for both RT,  $F(1, 33) = 23.03, p < .001, \eta_p^2 = .41$ , and error rates,  $F(1, 33) = 7.53, p = .01, \eta_p^2 = .19$ . This interaction was due to a larger Probe Type effect (Other-Oriented Probes minus Self-Oriented Probes) in perspective shift trials (153 ms; 3.53% errors) than in no perspective-shift trials (30 ms; 0.96% errors). In other words, within perspective-shift conditions, it was harder (taking longer,  $t(33) = -6.72, p < .001, r = .75$ , and with more errors,  $t(33) = -5.48, p < .001, r = .69$ ) to shift from Self-to-Other than from Other-to-Self (see Figure 2). In no perspective-shift conditions, there was no significant difference in RT,  $t(33) = -1.41, p = .17, r = .27$ , or error rates,  $t(33) = -1.61, p = .12, r = .32$ , between Self-to-Self trials and Other-to-Other trials (see Figure 2).

There was also a significant interaction between Perspective Shift and Contents Type for RT,  $F(1, 33) = 15.67, p = .001, \eta_p^2 = .32$ , and error rates,  $F(1, 33) = 4.20, p = .05, \eta_p^2 = .11$ . This was due to a larger effect of Contents Type on RT (Unexpected minus Expected) in no perspective shift conditions (164 ms), compared to perspective shift conditions (43 ms). For error rates, there was a larger effect of Contents Type in perspective shift conditions (6.4% errors) compared to no perspective shift conditions (3.89% errors).



**Figure 2:** Mean response time in milliseconds (top graph) and error rates (percentages; lower graph) for the different conditions of the Self/Other Differentiation Task, reflecting responses to the probe question. Error bars denote +/- 2 Standard Errors.

For error rates, there was a significant interaction between Contents Type and Probe Type,  $F(1, 33) = 29.15$ ,  $p < .001$ ,  $\eta_p^2 = .47$ , due to a larger effect of contents type (Unexpected minus Expected) for other-probes (7.09% errors) than for self-probes (3.20% errors).

Finally, there was a significant three-way interaction between Perspective Shift, Contents Type, and Probe Type for both RT,  $F(1, 33) = 8.64$ ,  $p = .006$ ,  $\eta_p^2 = .21$ , and error rates,  $F(1, 33) = 10.75$ ,  $p = .002$ ,  $\eta_p^2 = .25$ . Post-hoc analysis of RT suggested that this was because an interaction between perspective shift and probe type was only present for expected contents trials,  $F(1, 33) = 36.35$ ,  $p < .001$ ,  $\eta_p^2 = .52$ , but not for unexpected contents trials,  $F(1, 33) = .83$ ,  $p = .37$ ,  $\eta_p^2 = .03$ . In contrast, for error rates, this interaction between Perspective-Shifting and Probe-Type was present only in

unexpected contents conditions,  $F(1, 33) = 8.13$ ,  $p = .007$ ,  $\eta_p^2 = .20$ , and not for expected contents conditions,  $F(1, 33) = .81$ ,  $p = .37$ ,  $\eta_p^2 = .02$ . No other effects were significant (all  $p$  values  $> .05$ ).

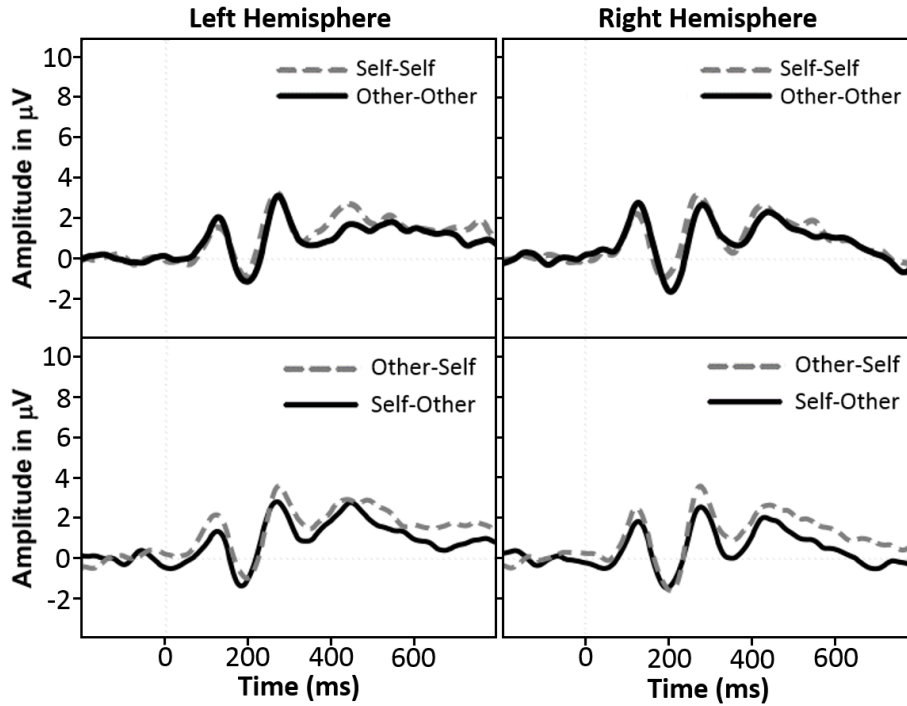
### ***Electrophysiological Data***

For analysis of the ERP data, only trials with correct responses were included. ERP analysis focused on two separate time points: (1) the onset of the probe question (Probe-Question locked ERPs) and (2) the onset of the probe answer options (Probe-Answer locked ERPs).

#### *Probe-Question Locked ERPs*

A significant interaction between Perspective-Shifting and Probe Type was present from 500 – 750ms after the presentation of the answer question across fronto-central areas,  $F_s(1, 33) \geq 4.40$ ,  $p_s \leq .044$ ,  $\eta_p^2s \geq .12$  (see Table 2 for details). This significant interaction was due to a larger negative difference in amplitude (for interval 600-650ms) between ‘No-Perspective Shift’ and ‘Perspective-Shift’ trials for ‘Other’ oriented probes ( $-0.91 \mu V$ ) than for ‘Self’ oriented probes ( $0.27 \mu V$ ).

A significant three-way interaction between Hemisphere, Perspective Shifting, and Probe Type was seen across parieto/occipito-lateral areas from 250 – 500 ms,  $F_s(1, 33) \geq 4.17$ ,  $p_s \leq .049$ ,  $\eta_p^2s \geq .11$  (see also Table 2). This was due to the significant interaction between Perspective Shifting and Probe Type being present only in the right hemisphere, but not in the left hemisphere. Figure 3 illustrates this effect.

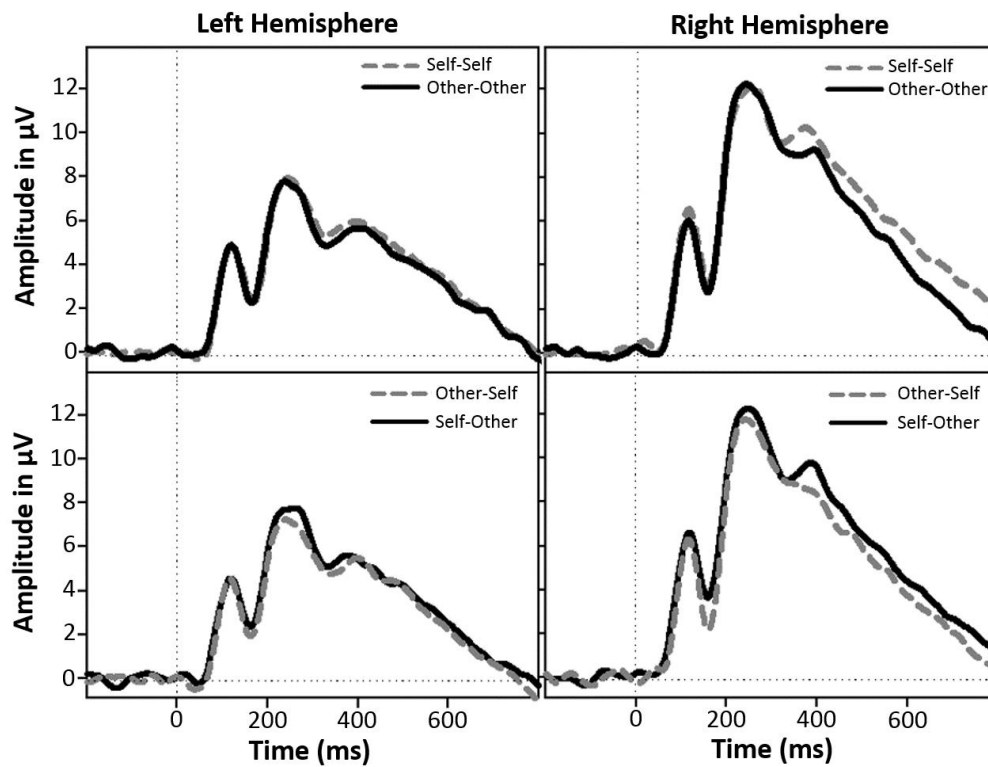


**Figure 3:** To show the significant Hemisphere\*Shift\*Probe interaction, ERP Waveforms are presented for the left hemisphere (left panels; electrode PO7) and separately for the right hemisphere (right panels; electrode PO8), with no perspective-shift trial conditions illustrated in the upper panels (Self-Self and Other-Other) and perspective-shift trial conditions illustrated in the lower panels (Other-Self and Self-Other), for Probe-Question Locked ERPs.

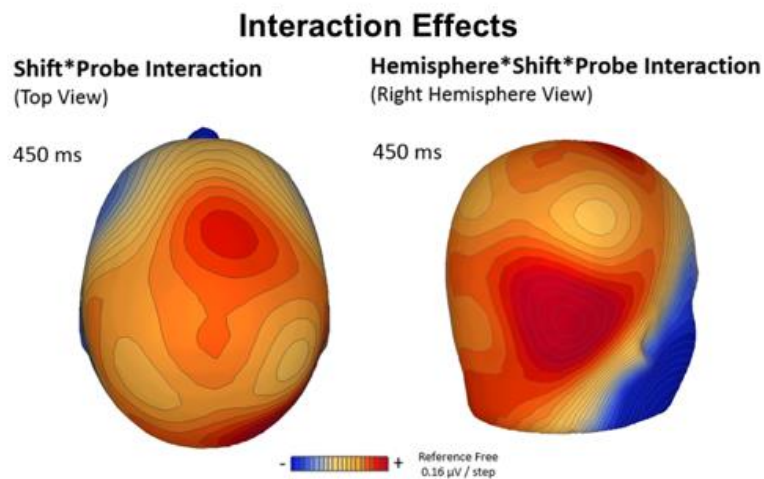
#### *Probe-Answer Locked ERPs*

A significant interaction between Perspective-Shifting and Probe Type was present from 350 – 500ms and 600 – 750ms after the presentation of the answer options across frontal-lateral areas,  $F_s(1, 33) \geq 4.32$ ,  $p_s \leq .046$ ,  $\eta_p^2s \geq .12$  (see Table 3 for details). This significant interaction was due to a larger difference in amplitude between ‘Self’ and ‘Other’ oriented probes in No-Perspective Shift conditions (0.47  $\mu V$ ) than in Perspective-Shift conditions (-0.23  $\mu V$ ). Figure 4 illustrates this effect.

A significant three-way interaction between Hemisphere, Perspective Shifting, and Probe Type was seen across parieto/occipito-lateral areas from 500 – 750 ms,  $F_s(1, 33) \geq 4.36$ ,  $p_s \leq .045$ ,  $\eta_p^2s \geq .12$  (see also Table 3). This was due to the significant interaction between Perspective Shifting and Probe Type being present only in the right hemisphere, but not in the left hemisphere (see Figures 4 and 5).



**Figure 4:** To show the significant Hemisphere\*Shift\*Probe interaction, ERP Waveforms are presented for the left hemisphere (left panels; electrode PO7) and separately for the right hemisphere (right panels; electrode PO8), with no perspective-shift trial conditions illustrated in the upper panels (Self-Self and Other-Other) and perspective-shift trial conditions illustrated in the lower panels (Other-Self and Self-Other), for Probe-Answer Locked ERPs.



**Figure 5:** Spline maps depicting the topographical distribution of the differences in amplitude [ $\mu\text{V}$ ] for the Shift\*Probe Interaction [(Self-Other minus Other-Self) minus (Other-Other minus Self-Self)], and the Hemisphere\*Shift\*Probe Interaction. These maps illustrate the spatial distribution of the interaction effect between Perspective-Shifting and Probe Type, for the Probe-Answer Locked ERPs.

**Table 2: Probe Question-locked ERP analysis:** Presentation of *F* values following Repeated-Measures ANOVAs of ERP Amplitude for the four pre-defined regions of interest. Note that, in an attempt to reduce the possibility of Type 1 errors, only results involving an interaction between Shift and Probe are reported here and only effects which were significant for three consecutive time windows were considered stable effects (see analysis section).

	Time from Onset of Probe Question in ms											
	150-200	200-250	250-300	300-350	350-400	400-450	450-500	500-550	550-600	600-650	650-700	700-750
<b>Fronto-Central Areas</b>												
Shift*Probe	1.65	1.51	2.36	<b>2.88(*)</b>	2.08	1.73	<b>3.81(*)</b>	<b>4.40*</b>	<b>6.99*</b>	<b>8.44**</b>	<b>8.35**</b>	<b>4.84*</b>
Shift*Contents*Probe	2.02	1.15	<b>4.69*</b>	<b>3.82(*)</b>	2.83	1.73	0.72	1.41	2.17	1.28	0.06	0.65
<b>Parieto-Central Areas</b>												
Shift*Probe	1.56	<b>11.37**</b>	2.13	1.76	0.09	0.02	0.03	0.00	0.25	0.36	1.68	1.15
Shift*Contents*Probe	0.08	0.22	0.33	0.68	0.03	0.44	0.82	1.49	1.29	0.82	1.85	0.44
<b>Fronto-Lateral Areas</b>												
Shift*Probe	0.00	0.21	0.36	0.07	0.13	0.21	1.28	1.89	2.09	<b>2.24(*)</b>	<b>5.96*</b>	<b>3.76(*)</b>
Hemi*Shift*Probe	0.72	0.00	0.01	0.21	0.15	0.68	0.03	1.25	0.51	0.38	0.29	1.14
Shift*Contents*Probe	0.17	1.61	1.13	0.02	0.07	0.19	0.46	1.04	0.31	0.18	0.43	0.08
Hemi*Shift*Contents*Probe	1.11	0.25	0.40	<b>3.38(*)</b>	2.29	<b>3.95(*)</b>	2.38	1.63	1.64	1.90	4.97*	<b>3.88(*)</b>
<b>Parieto/Occipito-Lateral Areas</b>												
Shift*Probe	0.39	<b>9.99**</b>	0.44	1.82	0.39	0.02	1.68	<b>3.60(*)</b>	<b>2.98(*)</b>	<b>5.88*</b>	<b>8.39**</b>	<b>3.41(*)</b>
Hemi*Shift*Probe	3.44(*)	<b>4.01(*)</b>	<b>4.60*</b>	<b>4.17*</b>	<b>9.01**</b>	<b>8.25**</b>	<b>6.20*</b>	0.40	2.02	1.72	2.61	2.52
Shift*Contents*Probe	0.16	2.15	1.00	0.44	1.69	1.08	1.52	0.87	1.91	0.38	1.61	0.90
Hemi*Shift*Contents*Probe	0.57	1.20	0.72	0.69	0.20	0.07	0.06	0.09	0.08	0.08	0.06	0.88

(\*)  $p < .10$ ; \*  $p < .05$ ; \*\* $p < .01$

**Table 3: Probe Answer-Options-locked ERP analysis:** Presentation of *F* values following Repeated-Measures ANOVAs of ERP Amplitude for the four pre-defined regions of interest. Note that, in an attempt to reduce the possibility of Type 1 errors, only results involving an interaction between Shift and Probe are reported here and only effects which were significant for three consecutive time windows were considered stable effects (see analysis section).

	Time from Onset of Probe Answer Options in ms											
	150-200	200-250	250-300	300-350	350-400	400-450	450-500	500-550	550-600	600-650	650-700	700-750
<b>Fronto-Central Areas</b>												
Shift*Probe	0.06	1.49	0.31	2.83	1.44	<b>2.10 (*)</b>	<b>4.36*</b>	0.88	0.98	1.21	<b>3.27 (*)</b>	2.05
Shift*Contents*Probe	0.10	0.46	0.01	1.23	0.24	0.26	0.01	0.75	<b>3.22 (*)</b>	<b>4.49*</b>	0.83	0.85
<b>Parieto-Central Areas</b>												
Shift*Probe	0.51	2.08	0.13	0.01	0.03	0.88	0.61	0.22	0.18	0.01	0.02	0.06
Shift*Contents*Probe	0.08	0.41	0.02	2.53	1.38	<b>3.14 (*)</b>	0.37	0.55	0.12	0.00	0.27	0.01
<b>Fronto-Lateral Areas</b>												
Shift*Probe	1.37	0.01	1.41	1.70	<b>4.32*</b>	<b>4.84*</b>	<b>4.52*</b>	<b>3.65 (*)</b>	1.83	<b>4.68*</b>	<b>5.79*</b>	<b>4.72*</b>
Hemi*Shift*Probe	0.00	0.08	0.40	0.02	0.04	0.02	0.17	0.00	0.04	0.02	0.04	0.14
Shift*Contents*Probe	0.00	0.00	0.00	0.07	0.03	0.89	0.11	0.12	0.01	0.12	0.44	0.36
Hemi*Shift*Contents* Probe	0.39	0.00	0.48	0.13	0.03	0.32	0.07	0.04	0.01	0.42	0.22	0.26
<b>Parieto/Occipito-Lateral Areas</b>												
Shift*Probe	<b>3.97(*)</b>	0.02	2.05	0.27	<b>3.27(*)</b>	1.64	1.69	2.22	1.45	2.21	2.17	3.82(*)
Hemi*Shift*Probe	1.83	<b>4.74*</b>	1.93	0.77	<b>3.39(*)</b>	<b>3.18(*)</b>	<b>3.00(*)</b>	<b>4.64*</b>	<b>4.36*</b>	<b>3.36(*)</b>	<b>5.08*</b>	<b>3.75(*)</b>
Shift*Contents*Probe	0.05	0.47	1.11	0.17	0.15	0.07	0.04	0.38	2.19	6.30	1.73	1.58
Hemi*Shift*Contents*Probe	1.78	<b>3.46(*)</b>	2.43	<b>3.92(*)</b>	0.64	0.71	0.39	1.53	1.19	0.21	0.00	0.07

(\*)  $p < .10$ ; \*  $p < .05$ ; \*\* $p < .01$



## DISCUSSION

This study utilized a matched-design false-belief task to allow direct comparison of ‘Self-Oriented’ and ‘Other-Oriented’ belief-attribution processes in ToM, as well as the role of Perspective-Shifting, at both a behavioural and electrophysiological level. The results of this study showed a distinct behavioural differentiation between the processes required for attribution of beliefs to oneself and other-people, with self-oriented belief attribution appearing to be much more efficient and successful (faster and less error prone) than other-oriented belief-attribution. Results also revealed a specific effect of ‘Perspective-Shifting’ in both behavioural outcomes and in the neurophysiological processes underlying this ability; trials in which no-perspective-shift was required were faster and more accurate than trials in which a perspective-shift was required. Critically, the *type* of perspective-shift required in a trial had a significant effect on responses; when a trial required a shift from Self-to-Other across the dilemma-to-probe, responses were significantly slower and more error prone than when a trial shifted from Other-to-Self. These behavioural effects replicate recent behavioural findings using the Self/Other Differentiation task (Bradford et al., 2015). The ERP results supported this perspective-shifting effect, suggesting a possible neurophysiological basis for the behavioural differences established; results showed that from 500ms post stimulus presentation a significant effect of Perspective-Shifting, interacting with Probe Type, was seen across fronto-central (probe-question locked ERPs) and fronto-lateral (probe-answer locked ERPs) areas. Additionally, a stable and consistent early-onset effect of a three-way interaction between Perspective-Shifting, Probe Type, and Hemisphere was seen from 250ms post stimulus presentation, suggesting lateralisation of these effects, with the critical Perspective-Shift x Probe Type interaction only being present in the right hemisphere. Results of the current study suggest that, supporting prior research, areas across both frontal and parietal areas of the brain may form core components of the ToM ‘network’, working together to engage in ToM processes, and playing an important role in coordination of ‘Self’ and ‘Other’ perspectives.

The results of this study support prior literature that suggests that one's own mental states may be processed separately from computation and assessment of the mental states of other people (e.g., Frith & Frith, 2003; Leslie & Thaiss, 1992; Perner et al., 2006). It appears that self-oriented and other-oriented belief-attribution may recruit separable and distinguishable processes, albeit likely with some overlap between them. In line with Vogeley et al. (2001) and Epley et al. (2004), the current results also indicate that the 'Self' perspective may form the prerequisite stem for understanding the perspectives of 'Others'. In Self-to-Other switch trials, priming of 'Self' at the dilemma stage meant that the 'Other' perspective was not computed until explicitly required (probe stage), increasing response times and the likelihood of making errors. In contrast, in Other-to-Self switch trials, participants efficiently (faster and more accurately) switched from the primed 'other' perspective to the 'self' perspective, suggesting that the 'self' may have been processed as a way of understanding the 'other'. Thus, the 'other-perspective' is processed only when explicitly required, whilst the 'self-perspective' is processed regardless of ultimate task demands. Further supporting this, in no perspective-shift trials, there was no significant difference in response times or error rates to either 'Self' or 'Other' oriented probe questions. The perspective primed at the dilemma stage was upheld at the probe stage, allowing preparation of responses, maintained across a trial. These findings indicate a specific performance effect of perspective-shifting, as also indicated by Ferguson et al. (2017) and Samson et al. (2010), and are further supported by the significant interaction between Perspective-Shifting and Probe Type revealed in our ERP analysis.

Analysis of ERP results was time-locked to two separate stimulus-onset points: the onset of the probe-question, and the onset of the probe answer-option images. Results from the ERPs time-locked to the probe question presentation demonstrated an early onset three-way interaction between Perspective-Shifting, Probe Type, and Hemisphere, first seen from 250ms post-stimulus onset, across parieto/occipito-lateral areas. Additionally, a two-way interaction between Perspective-Shifting and Probe Type was seen from 500ms post-stimulus presentation, maximal over fronto-central areas of the brain. Results from the ERPs time-locked to the probe answer-options

presentation highlighted a significant two-way interaction between Perspective-Shifting and Probe Type from 350ms post-stimulus onset, across fronto-lateral areas, with a significant three-way interaction between Perspective-Shifting, Probe Type and Hemisphere seen across parieto/occipito-lateral areas, from 500ms post-stimulus onset (probe answer-options). These results are in line with prior research that has indicated the importance of both frontal regions and parietal regions in processing of ToM (e.g. Frith & Frith, 2003; Happé, 2003; Sabbagh & Taylor, 2000; Saxe & Wexler, 2005; Gallagher et al., 2000; Schurz et al., 2014). Results indicate that the requirement of shifting between perspectives, and the direction of this shift – orienting to the ‘self’ or ‘other’ perspective – is processed at a relatively early onset and, critically, that this consideration continues to be processed until the resolution of a trial.

The current results suggest that, in a ToM task requiring participants to attribute beliefs to both themselves and other people, key neural activity can be seen across a network of regions, including frontal-lateral and parieto/occipito-lateral regions. Previous research has indicated that this network, including the anterior cingulate cortex, the pre-frontal cortex, and the right temporal parietal junction, may play key roles in managing and reconciling contrasting perspectives, as well as being involved in switching between ‘Self’ and ‘Other’ perspectives (e.g., Amodio & Frith, 2006; Gallagher et al., 2000; Gobbini et al., 2007; Saxe & Kanwisher, 2003; McCleery et al., 2011; Saxe & Powell, 2006; Schurz et al., 2014). It could be suggested that in the current study, ERP activation found over frontal areas may reflect activation of the prefrontal cortex, due to the need to process conflicting perspectives of the self and other, and consideration of whether a need for a switch in perspective is required to successfully resolve and answer a given probe-perspective question (e.g. Gobbini et al., 2007; Gallagher et al., 2000; Frith & Frith, 2003; Saxe et al., 2006; Samson et al., 2005). The results of the current study revealed a three-way interaction between Perspective-Shifting, Probe-Type and Hemisphere, which may be indicative of the processing required to adapt to the necessary perspective prompted, and to adopt the relevant perspective for successfully answering the probe question, with activity localised to the right hemisphere (e.g. Saxe & Powell,

2006; Saxe & Wexler, 2005; McCleery et al., 2011; Schurz et al., 2014; Gallagher & Frith, 2003). Prior studies have suggested brain areas, such as the temporo-parietal cortex and posterior cingulate cortex, may be required for the disengagement of the 'Self' perspective and the adoption of the 'Other' perspective, supporting the current findings (e.g. Happé, 2003; Schurz et al., 2014; McCleery et al., 2011; Gobbini et al., 2007); however, due to the limited spatial resolution of ERP measures, identification of potential underlying brain areas in the current study are interpreted with caution.

As discussed, the results of the current research both support and expand on prior literature. Of key importance, this study allowed direct comparison of 'Self' and 'Other' oriented belief-attribution processes, as well as the process of shifting between these two perspectives, at a behavioural and electrophysiological level. The use of ERP measures allowed exploratory examination of broad measures of neural activity associated with the behavioural effects ascertained, and the time-course of these effects. Results revealed a key role of perspective-shifting, interacting with both probe type and hemisphere, over frontal-lateral and parieto/occipito-lateral areas, supporting prior research findings that have implicated, for example, the temporo-parietal junction and the anterior cingulate cortex as playing prominent roles in ToM processes (e.g. Happé, 2003; Saxe & Kanwisher, 2003; Gobbini et al., 2007; Saxe & Wexler, 2005; Vogeley et al., 2001; Gallagher & Frith, 2003). These results highlight some key focus areas for further exploration, to assess the independent, individual roles differential neural components may be playing in ToM expression. One of the key manipulations in the Self/Other Differentiation task was the creation of matched-trials, in which the answer to either self-oriented or other-oriented belief attribution trials was the same (i.e. both 'Self' and 'Other' believe there to be sugar in the sugar jar). In this way, any differentiation in the neural activity associated with self-oriented, other-oriented and perspective-shift trials appears to reflect the processes undertaken to adopt a specific perspective, or to engage, maintain or switch to a required perspective, rather than reflecting calculations of different solutions to probe questions. The solution is the same in both 'self' and 'other' probe questions, suggesting any calculations undertaken reflect alignment to the appropriate response perspective, of either self or other.

### *Conclusion*

A key contribution of the current study is the opportunity to directly compare ‘Self’ and ‘Other’ belief-attribution processes, and to examine the specific role of perspective-shifting in ToM capacities, using a matched design. Results indicated that ‘Self’ and ‘Other’ belief-attribution processes, a key part of the ToM mechanism, can be distinguished at a behavioural level (response-time/accuracy) within the same task. Crucially, the results provide evidence of ‘Self’ and ‘Other’ differentiation processing in the ToM capacity, and the specific role of the process of ‘Perspective-Shifting’, at both a behavioural and neural (ERP measures) level. Perspective-shifting is hypothesized to play a key role in everyday interaction and communication, when decisions about what to say or how to act must frequently take into account the contrast between what others know and what we ourselves know. The interaction between Perspective-Shifting and Probe Type was found over both frontal-lateral and right parieto/occipito-lateral areas, suggesting a possible role of the involved brain areas in the assessment and calculation of whether it is necessary to adopt the ‘Self’ or ‘Other’ perspective. The right lateralization of the parieto/occipito-lateral effect also suggests a potentially specific role in ToM processing, and is supported by prior research that has indicated, for instance, the role of the right temporo-parietal junction in ToM engagement (e.g. Saxe & Powell, 2006; McCleery et al., 2011; Schurz et al., 2014). The current task provides some indication of the more specific role that these components may be playing in ToM processing, such as the explicit role of perspective-shifting, as well as that of differentiating between ‘Self’ and ‘Other’. Our results, obtained with a typically developed, healthy adult population, provide the opportunity for further study into how these processes unfold during ToM development and how they may atypically develop in conditions such as autism (Baron-Cohen et al., 2001; Brent et al., 2004).

### **AUTHOR CONTRIBUTIONS**

All authors contributed to the study concept and design. Testing and data collection were performed by EB. EB performed the data analysis and interpretation under the supervision and help from JG

and IJ. EB drafted the manuscript, and JG and IJ provided critical revisions and comments. All authors approved the final manuscript.

#### **COMPETING INTERESTS**

The authors declare no competing interests.

#### **ACKNOWLEDGEMENTS**

This work was supported by the Economic and Social Research Council [grant number ES/J500136/1] in the form of a three-year PhD studentship awarded to EB.

## REFERENCES

- Amodio, D.M., & Frith, D.F. (2006). Meeting of the Minds: the Medial Frontal Cortex and Social Cognition. *Nature Reviews Neuroscience*, 7, 268-277. doi: 10.1038/nrn1884
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The 'Reading the Mind in the Eyes' Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-Functioning Autism. *Journal of Child Psychology and Psychiatry*, 42, 241-251. doi: 10.1111/1469-7610.00715
- Bradford, E.E.F., Jentsch, I., & Gomez, J-C. (2015). From Self to Social Cognition: Theory of Mind Mechanisms and their Relation to Executive Functioning. *Cognition*, 138, 21-34. doi: 10.1016/j.cognition.2015.02.001
- Bradford, E.E.F., Jentsch, I., Gomez, J-C., Chen, Y., Zhang, D., & Su, Y. (2018). Cross-cultural differences in adult Theory of Mind abilities: A comparison of native-English speakers and native-Chinese speakers on the Self/Other Differentiation task. *Quarterly Journal of Experimental Psychology*, 1-12. doi: 10.1177/1747021818757170
- Brent, E., Rios, P., Happé, F., & Charman, T. (2004). Performance of Children with Autism Spectrum Disorder on Advanced Theory of Mind Tasks. *Autism*, 8, 283-299. doi: 10.1177/1362361304045217
- Campbell, T., Kerlin, J.R., Bishop, C.W., & Miller, L.M. (2012). Methods to eliminate stimulus transduction artifact from insert earphones during electroencephalography. *Ear and Hearing*, 33(1), 144-150. doi: 10.1097/AUD.0b013e3182280353
- Cane, J.E., Ferguson, H.J., & Apperly, I. (2017). Using Perspective to Resolve Reference: the impact of cognitive load and motivation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(4), 591-610. doi: 10.1037/xlm0000345
- Catmur, C., Santiesteban, I., Conway, J.R., Heyes, C., & Bird, G. (2016). Avatars and arrows in the brain. *NeuroImage*, 132, 8-10.
- Conway, J.R., Lee, D., Ojaghi, M., Catmur, C., & Bird, G. (2017). Submentalizing or mentalizing in a Level 1 perspective-taking task: A cloak and goggles test. *Journal of Experimental Psychology: Human Perception and Performance*, 43(3), 454-465.
- Decety, J., & Sommerville, J.A. (2003). Shared Representations Between Self and Other: a Social Cognitive Neuroscience View. *TRENDS in Cognitive Sciences*, 7, 527-533. doi: 10.1016/j.tics.2003.10.004
- Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of personality and social psychology*, 87(3), 327-339
- Ferguson, H.J., Apperly, I., & Cane, J.E. (2017). Eye-tracking reveals the cost of switching between self and other perspectives in a visual perspective-taking task. *Quarterly Journal of Experimental Psychology*, 70(8), 1646-1660. doi: 10.1080/17470218.2016.1199716

- Ferguson, H.J., Cane, J.E., Douchkov, M., & Wright, D. (2015). Empathy predicts false belief reasoning ability: evidence from the N400. *Social, Cognitive and Affective Neuroscience*, 10(6), 848-855. doi:10.1093/scan/nsu131
- Frith, U., & Frith, C.D. (2003). Development and Neurophysiology of Mentalizing. *Philosophical Transactions of the Royal Society of London (Series B)*, 358, 459-473. doi: 10.1098/rstb.2002.1218
- Gallagher, H.L., & Frith, C.D. (2003). Functional Imaging of 'Theory of Mind'. *TRENDS in Cognitive Science*, 7(2), 77-83. doi: 10.1016/S1364-6613(02)00025-6
- Gallagher, H.L., Happé, F., Brunswick, N., Fletcher, P.C., Frith, U., & Frith, C.D. (2000). Reading the Mind in Cartoons and Stories: an fMRI Study of 'Theory of Mind' in Verbal and Nonverbal Tasks. *Neuropsychologia*, 38, 11-21. doi: 10.1016/S0028-3932(99)00053-6
- Gallagher, S., & Zahavi, D. (2012). *Phenomenological Mind*. London, UK: Routledge
- Gallese, V. (2006). Intentional Attunement: A neurophysiological perspective on social cognition and its disruption in autism. *Brain Research*, 1079(1), 15-24. doi: 10.1016/j.brainres.2006.01.054
- Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *TRENDS in Cognitive Sciences*, 8(9), 396-403. doi: 10.1016/j.tics.2004.07.002
- Geng, J.J., & Vossel, S. (2013). Re-evaluating the role of the TPJ in attentional control: Contextual updating? *Neuroscience & Biobehavioral Reviews*, 37, 2608-2620.
- Gobbini, M., Koralek, A., Bryan, R., Montgomery, K., & Haxby, J. (2007). Two Takes on the Social Brain: A Comparison of Theory of Mind Tasks. *Journal of Cognitive Neuroscience*, 19(11), 1803-1814. doi: 10.1162/jocn.2007.19.11.1803
- Gopnik, A., & Astington, J.W. (1988). Children's Understanding of Representational Change and its Relation to the Understanding of False-Belief and the Appearance-Reality Distinction. *Child Development*, 59, 26-37. doi: 10.2307/1130386
- Grimm, S., Roeber, U., Trujillo-Barreto, N.J., & Schröger, E. (2006). Mechanisms for detecting auditory temporal and spectral deviations operate over similar time windows but are divided differently between the two hemispheres. *NeuroImage*, 32(1), 275-282. doi: 10.1016/j.neuroimage.2006.02.032
- Happé, F. (2003). Theory of Mind and the Self. *Annals New York Academy of Sciences*, 1001, 134-144. doi: 10.1196/annals.1279.008
- Hartwright, C.E., Apperly, I.A., & Hansen, P.C. (2012). Multiple Roles for Executive Control in Belief-Desire Reasoning: Distinct Neural Networks are Recruited for Self Perspective Inhibition and Complexity of Reasoning. *NeuroImage*, 61, 921-930. doi: 10.1016/j.neuroimage.2012.03.012
- Hogrefe, G.J., Wimmer, H., & Perner, J. (1986). Ignorance versus False-Belief: A Developmental Lag in Attribution of Epistemic States. *Child Development*, 57(3), 567-582. doi: 10.2307/1130337
- Ille, N., Berg, P., & Scherg, M. (2002). Artifact Correction of the Ongoing EEG Using Spatial Filters based on Artifact and Brain Signal Topographies. *Journal of Clinical Neurophysiology*, 19, 113-124. doi: 10.1097/00004691-200203000-00002



- Keysar, B., Barr, D.J., Balin, J.A., & Brauner, J.S. (2000). Taking Perspective in Conversation: the Role of Mutual Knowledge in Comprehension. *Psychological Science*, 11(1), 32-38. doi: 10.1111/1467-9280.00211
- Leslie, A.M., & Thaiss, L. (1992). Domain Specificity in Conceptual Development: Neuropsychological Evidence from Autism. *Cognition*, 43(2), 225-251. doi: 10.1016/0010-0277(92)90013-8
- Liu, D., Sabbagh, M.A., Gehring, W.J., & Wellman, H.M. (2009). Neural Correlates of Children's Theory of Mind Development. *Child Development*, 80(2), 318-326
- Mahy, C.E.V., Moses, L.J., & Pfeifer, J.H. (2014). How and Where: Theory-of-Mind in the Brain. *Developmental Cognitive Neuroscience*, 9, 68-81. doi: 10.1016/j.dcn.2014.01.002
- McCleery, J.P., Surtees, A.D.R., Graham, K.A., Richards, J.E., & Apperly, I.A. (2011). The Neural and Cognitive Time Course of Theory of Mind. *The Journal of Neuroscience*, 31(36), 12849-12854. doi: 10.1523/JNEUROSCI.1392-11.2011
- Meinhardt, J., Sodian, B., Thoermer, C., Dohnel, K., & Sommer, M. (2011). True- and false-belief reasoning in children and adults: An event-related potential study of theory of mind. *Developmental Cognitive Neuroscience*, 67-76
- Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron*, 50, 655-663. doi: 10.1016/j.neuron.2006.03.040
- Perner, J., Aichhorn, M., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Thinking of Mental and Other Representations: the Roles of Left and Right Temporo-Parietal Junction. *Social Neuroscience*, 1(3-4), 245-258. doi: 10.1080/17470910600989896.
- Premack, D., & Woodruff, G. (1978). Does the Chimpanzee have a Theory of Mind? *Behavioural and Brain Sciences*, 1, 515-526. doi: 10.1017/S0140525X00076512
- Sabbagh, M.A., & Taylor, M. (2000). Neural Correlates of Theory of Mind Reasoning: An Event-Related Potential Study. *Psychological Science*, 11(1), 46-50. doi: 10.1111/1467-9280.00213
- Sabbagh, M.A., Bowman, L.C., Evraire, L.E., & Ito, J.M.B. (2009). Neurodevelopmental Correlates of Theory of Mind in Preschool Children. *Child Development*, 80(4), 1147-1162
- Samson, D., Apperly, I.A., Braithwaite, J.J., & Andrews, B.J. (2010). Seeing it Their Way: Evidence for Rapid and Involuntary Computation of What Other People See. *Journal of Experimental Psychology: Human Perception and Performance*, 36, 1255-1266. doi: 10.1037/a0018729
- Samson, D., Apperly, I.A., Kathirgamanathan, U., & Humphreys, G.W. (2005). Seeing it My Way: A Case of a Selective Deficit in Inhibiting Self-Perspective. *Brain*, 128, 1102-1111. doi: 10.1093/brain/awh464
- Santiesteban, I., Catmur, C., Hopkins, S.C., Bird, G., & Heyes, C. (2014). Avatars and arrows: Implicit mentalizing or domain-general processing? *Journal of Experimental Psychology: Human Perception and Performance*, 40(3), 929-937

- Saxe, R., & Kanwisher, N. (2003). People Thinking About Thinking People: The Role of the Temporo-Parietal Junction in 'Theory of Mind'. *NeuroImage*, 19(4), 1835-1842. doi: 10.1016/S1053-8119(03)00230-1
- Saxe, R., & Powell, L.J. (2006). It's the Thought That Counts: Specific Brain Regions for One Component of Theory of Mind. *Psychological Science*, 17(8), 692-699. doi: 10.1111/j.1467-9280.2006.01768.x
- Saxe, R., & Wexler, A. (2005). Making Sense of Another Mind: the Role of the Right Temporo-Parietal Junction. *Neuropsychologia*, 43(10), 1391-1399. doi: 10.1016/j.neuropsychologia.2005.02.013
- Saxe, R., Moran, J.M., Scholz, J., & Gabrieli, J. (2006). Overlapping and Non-Overlapping Brain Regions for Theory of Mind and Self Reflection in Individual Subjects. *Social Cognitive and Affective Neuroscience*, 1, 229-234. doi: 10.1093/scan/nsl034
- Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating Theory of Mind: A Meta-Analysis of Functional Brain Imaging Studies. *Neuroscience and Biobehavioral Reviews*, 42, 9-34. doi: 10.1016/j.neubiorev.2014.01.009
- Saxe, R., Whitfield-Gabrieli, S., Scholz, J., & Pelphrey, K.A. (2009). Brain Regions for Perceiving and Reasoning About Other People in School-Aged Children. *Child Development*, 80(4), 1197-1209. doi: 10.1111/j.1467-8624.2009.01325.x.
- Sysoeva, O. V., Lange, E. B., Sorokin, A. B., & Campbell, T. (2015). From pre-attentive processes to durable representation: An ERP index of visual distraction. *International Journal of Psychophysiology*, 95, 310-321. doi: 10.1016/j.ijpsycho.2014.12.007
- Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happé, F., Falkai, P., Maier, W., Shah, N.J., Fink, G.R., & Zilles, K. (2001). Mind Reading: Neural Mechanisms of Theory of Mind and Self-Perspective. *NeuroImage*, 14, 170-181. doi: 10.1006/nimg.2001.0789
- Williams, D., & Happé, F. (2009). What Did I Say? Versus What Did I Think? Attributing False-Beliefs to Self Amongst Children With and Without Autism. *Journal of Autism and Developmental Disorders*, 39, 865-873. doi: 10.1007/s10803-009-0695-6
- Wimmer, H., & Perner, J. (1983). Beliefs About Beliefs: Representation and Constraining Function of Wrong Beliefs in Young Children's Understanding of Deception. *Cognition*, 13, 103-128. doi: 10.1016/0010-0277(83)90004-5